ACIG
APPLIED CYBERSECURITY & INTERNET GOVERNANCE

NASK

Trust Framework on Exploitation of Humans as the Weakest Link in Cybersecurity

Morice Daudi | Department of Computing Science Studies, Faculty of Science and Technology, Mzumbe University, United Republic of Tanzania,
ORCID: 0000-0001-7907-427X

Abstract

The significance of cybersecurity is increasing in our daily digital lives. The reason for this rise is that human interactions take place in computer-mediated environments, or cyberspace, where physical cues from face-to-face interactions are either absent or very minimal. Computer users are becoming increasingly susceptible to cyberattacks as a result of human interactions in cyberspace. Understanding how cybercriminals exploit the human trust, the weakest link in cybersecurity is relevant because cybercriminals focus on attacking the human psychology of trust rather than technical-based controls. To this end, the present paper develops a trust framework on exploitation of humans as the weakest link in cybersecurity. The framework is established by linking the human psychology of trust and techniques used by cybercriminals in deceiving and manipulating users of computer systems. The framework is validated by demonstrating its application using a case study employing real data. Findings show that cybercriminals exploit human trust based on trust development processes and bases of trust, either creating (falsified) expectations or a relationship history to lure the victim in. Furthermore, it is revealed that technical-based controls cannot provide effective safeguards to prevent manipulation of the human psychology of trust.

Received: 29.06.2023

Accepted: 16.10.2023

Published: 31.12.2023

Cite this article as:
M. Daudi "Trust Framework on Exploitation of Humans as the Weakest Link in Cybersecurity," ACIG, vol. 2, no. 1, 2023, doi: 10.60097/ACIG/162867.

Corresponding author:
Morice Daudi, Department of Computing Science Studies, Faculty of Science and Technology, Mzumbe University, United Republic of Tanzania,
E-MAIL:
morice.daudi@mu.ac.tz;
dmorice@mzumbe.ac.tz

Keywords
cybersecurity, human layer, weakest link, trust, trust framework, human trust exploitation

1. Introduction

Cyberattacks have evolved in many forms and stages. From attacking computers and computer networks, today's cyberattacks also target human beings. This progression, which Schneier [1] refers to as waves of attacks, consists of physical attacks, attacks that target vulnerabilities, and semantic attacks. According to the author, the first wave comprises attacks against computers, wires, and electronics, the second targets vulnerabilities in software products, cryptographic algorithms, protocols, and denial-of-service, and the third targets how humans assign meaning to content. State officials in charge of upholding the law, such as the police, have documented numerous incidences, particularly for the third wave. For example, between 2017 and 2020, 19,530 cybercrime incidents were reported to police in Tanzanian [2–5]. Many of these incidents were committed through social engineering techniques relating to how humans assign meaning to content. Such ever-increasing cybercrimes are emphasized by Schneier [1], who posits that semantic attacks will become more serious than physical or even syntactic attacks in future and that dismissing them using cryptographic measures will be difficult. Given that humans are the weakest link in computer information system components, semantic attacks target people more than other components. That shift in target is partly attributed to the relative strength of technical-based controls in cybersecurity. Technical-based controls are difficult to crack compared to human being psychology, which is easy to manipulate.

Technical controls are built on the triad of Confidentiality, Integrity, and Availability (CIA). These principles are widely used to ensure the security of computer resources. Despite their advantages, CIA concepts have several drawbacks. First, the CIA framework focuses on isolating legitimate from illegitimate users, granting legitimate users full access to computer resources, which that user is privileged to access. Once users are considered fair and granted access, CIA primitives provide the least control over actions users can perform. Second, CIA principles rely on algorithms developed based on historically conceptualised cybersecurity incidents. That history dependence implies that new incidents that have not been conceived are difficult to control and manage. Given these restrictions and a rise in attacks on humans compared to cryptograph-based

methods, there is a need for human-centric complementary defence. That need is imperative because technology is not the only way to address information security risks [6]. Furthermore, customers and organisational insiders make information security challenging [7], as their misbehaviour can directly or indirectly lead to cybercrime. Since most amateurs attack machines while professionals target people, cybersecurity solutions must now target humans more [1] than ever before. Creating human-centric cybersecurity solutions necessitates collaboration between industry and academia to brainstorm from an alternative perspective. One of those perspectives is trust, a human component many cybercriminals exploit. The critical question may be, "How do humans come to trust cybercriminals?"

Humans play trusting roles in cybersecurity at a moment when technical-based controls fail to detect and prevent cyberattacks. One area contributing to cybercrime attacks involves trust between computer users and cybercriminals. Cybercriminals prefer to exploit people's trust rather than technology since it is easier to exploit their natural inclination to trust [8]. It is also simple to deceive people if you can gain their trust. For this reason, cybersecurity requires managerial efforts on top of technical-based controls to combat cybercrime.

Cybercrime occurs at many layers, much like those of Open Systems Interconnection (OSI) and TCP/IP models. These cybersecurity layers include mission critical assets, data security, application security, endpoint security, network security, perimeter security, and the human layer [9, 10]. Among them, human is the weakest and most vulnerable layer. The co-existence of these layers implies that technological and management or policy-based control must be implemented nearly concurrently. Though they coexist, the human layer depends more on policy-based controls than technical-based solutions. This is due to the fact that human trust behaviours manifest in reasoned decisions and actions that are not part of coded algorithms. Rather, trust behaviour results from an individual's mental ability to either accept or reject cooperation with a counterpart based on the degree of trust the user builds. Trust in humans is attributed to inherent characteristics, which are part of the individual or "given" by the trust-giver, and situational characteristics external to the individual [11].

Many technical solutions have been developed to counteract cyberattacks. However, the number of cyberattacks continues to increase due to inherent constraints in CIA doctrines and a shift in emphasis on exploiting humans as the weakest link. In [12] the authors

describe technical and non-technical state-of-the-art protection tools related to everyday online activities. In [13], the authors analyse models of human behaviour that impact data system protection and how systems can be improved and highly secured against any vulnerabilities. These options, which are typical of numerous existing alternatives, are insufficient on their own. Human beings continue to be the weakest link in cybersecurity, a fact that cybercriminals know and take advantage of by exploiting the human psychology of trust. Meanwhile, algorithms for detecting and preventing human trust exploitation are scarce in the literature. This is confirmed by Shabut et al., who contend that an intelligent tool capable of comprehending cyberattack mechanisms and user behaviours involving assumptions, decision-making, and responses to cyber threats/risks is currently lacking [12]. Alternatively, users must be aware of how cybercriminals exploit human trust instead of depending only on coded algorithms. To this end, the overall objective of the present paper is to examine in detail how cybercriminals exploit human trust. The paper contributes by formulating a trust framework on the exploitation of humans as the weakest link in cybersecurity. It answers the two following research questions:

- How do cybercriminals exploit human trust, the most vulnerable link in cybersecurity?
- How can the development of a trust framework on the exploitation of humans as the weakest link in cybersecurity be beneficial?

The paper contributes to helping individuals and organisations know and gain awareness of trust development processes and bases of trust that cybercriminals employ to manipulate and deceive users of digital systems and gadgets. That awareness enables individuals and organisations to detect, react and prevent attacks on human trust, leaving them better equipped to recognise, respond to, and stop such attacks.

2. Trust and Cybersecurity in the Human Layer

The present section covers discussions on trust and cybersecurity in the human layer. The human layer of cybersecurity is covered in subsection 2.1. Common cyberattacks affecting individuals and organisations are covered in more detail in subsection 2.2. Subsection 2.3 concludes the discussion by presenting a thorough analysis of trust in computer-mediated environments.

2.1. Cybersecurity in the Human Layer

The human layer of cybersecurity is part of cyberspace, a time-dependent set of interconnected information systems and human users that interact with these systems [14]. It is in this space, cyberspace, where cybercrime occurs. Cybercrime can essentially be regarded as any crime (traditional or new) that can be conducted or enabled through digital technologies [15]. Such crimes must be controlled and prevented to safeguard data, information systems, and users. Consequently, the act of detecting, reacting and preventing cybercrime is referred to as cybersecurity. The authors in [16] define cybersecurity as the organisation and collection of resources, processes, and structures used to protect cyberspace and cyberspace-enabled systems from occurrences that misalign de jure from de facto property rights. With that brief overview, the next paragraph contextualises cybersecurity in the human layer.

Today's cryptographic magic wands of "digital signatures", "authentication", or "integrity" [1] are not the ultimate protective mechanism to rely on. These cryptographic techniques can barely identify most lies that manipulate the human psychology of trust. Cybercriminals have a long history of taking advantage of the psychological needs and vulnerabilities of people in a variety of ways, including the human need for love and affection, our fundamental desire to be trustworthy and helpful, and the many biases that influence security decision-making [17]. Another form relates to perfect knowledge of what people consider most important [15]. These outlined techniques are sources of human weaknesses that cybercriminals employ as weapons to exploit individuals and organisations.

2.2. Cyberattacks in Cyberspace

Cybersecurity incidents impact individuals and organisations worldwide, causing harm to social and economic values. They involve malware, password theft, traffic interception, phishing, denial-of-service, cross-site (xss), zero-day exploits, social engineering, and crypto-jacking. However, other types of cybercrime, such as terrorism, cyber warfare, cyber espionage, and cyberbullying, are also emerging. All of these threats originate in the digital environment in networked and non-networked computer systems.

Cybercrime threats affect our everyday life, from financial transactions to social interactions. For example, reports from the Inspector General of Police show that over four years, 19,550 incidents of cybercrime were reported in Tanzania [2–5] (Table 1).

Table 1. Selected cybersecurity incidents in four years in Tanzania [2–5].

Type of Cyber Incident	Year				Cumulative Sum
	2017	2018	2019	2020	
Theft	2,568	4,310	2,408	2,963	12,249
Death Threats	447	851	409	490	2,197
Insults	489	757	287	385	1,918
Threats	51	61	43	254	409
Misuse of the Internet	11	374	19	0	404
Trusted Theft	81	203	31	65	380
Fraud	48	282	1	21	352
ATM Theft	171	20	97	53	341
Financial Fraud	52	40	83	89	264
Forgery	78	112	21	32	243
Attempted Financial Fraud	17	2	10	111	140
Network and System Intrusion	26	15	0	54	95

According to that report, theft, death threats, and insults are the major cybercrime incidents being reported to the police. Such statistics correspond to a remark emphasised in The Citizen that theft via mobile money transactions, abusive language, and theft of information shared on various cyber platforms are frequently committed crimes [18]. It is estimated that 91% of cybercrime cases go unreported to the police [19], suggesting that 19,530 cybercrime recorded incidents may reflect underreporting of cases by organisations and individuals.

Some national and international organisations are already implementing strategies to fight cybercrime. For instance, AFRIPOL [20]has been fighting cybercrime by raising awareness, reinforcing policy and legislation to fight cyber criminals, and implementing technologies to support cyber-defence. Similar measures to combat cybercrime are also recommended in other literature sources. The Tanzania Cybersecurity Report of 2016 recommends improving internet user

education [19] in fighting cybercrime. Educating users also means raising their cybersecurity awareness, which is critical, especially for organisations that have many employees. Since research shows that over 80% cases of system-related fraud and theft in 2016 were perpetrated by employees and other insiders [19], training employees on proper internet use and how to fight cyberattacks is essential. Moreover, it is indispensable to extend training on cybersecurity awareness to individuals in the local community.

Conversely, cybercriminals play on human psychology to manipulate users, and gain or guess their access credentials. Evidence of this claim is featured in weekly reports[1] released by [21] TZ-CERT in Tanzania. TZ-CERT studies cyberattack patterns by setting up a honeypot. The honeypot is a network-attached system set up as a decoy to lure cyber attackers and detect, deflect or study hacking attempts meant to gain unauthorised access to information systems. The resulting information helps to guide users of computer systems in many ways, including how to prevent cyberattacks. According to TZ-CERT reports, which are analysed in Figure 1, cybercriminals use human psychological heuristics – based on the human inclination to use default, simple, or common access credentials – to guess usernames and passwords.

1 —— Seven reports released from 5 to 17 July 2023.

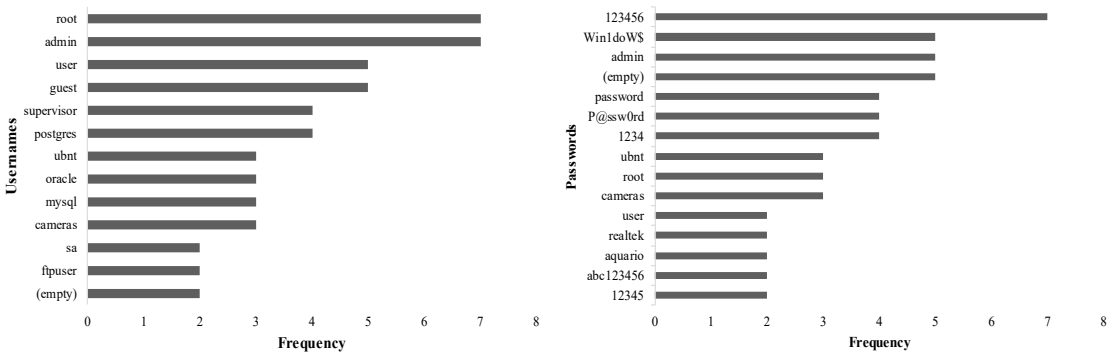

Figure 1. Common usernames and passwords used by cybercriminals.

Usernames such as "*root*", "*admin*", "*user*", "*guest*", "*supervisor*", and "*postgres*", and passwords such as "*123456*", "*win1doW$*", "*admin*", "*(empty)*" and "*password*" were common. These usernames and passwords are easy to remember, hence their prevalence. The access credentials in Figure 1 presents the psychological behaviour of many computer system users when choosing usernames and passwords.

2.3. Trust in Computer-Mediated Environment

The current section discusses trust in a computer-mediated setting, investigating the risk viewpoint of trust (subsection 2.3.1), as well as behavioural control in interactive systems (subsection 2.3.2). It concludes with a discussion of bases of trust (subsection 2.3.3).

2.3.1. Risk Perspective on Trust in Relationships

For an exchange to be completed, two parties, a trustor and a trustee, must be engaged. A trustor is an entity that develops a degree of reliance on another object and accepts being vulnerable to the possible actions of that other object [22]. Similarly, the trustee is the party in whom the trust resides, who can exploit the trustor's vulnerabilities [23]. The trustor is the party that puts its expectations in the other party, while the trustee is the party in which that expectation resides. While many definitions of trust exist, the present paper adopts the following definition: Trust is a level of confidence a trustor develops in a trustee based on the expectation that the trustee will perform a particular action necessary to the trustor [22].

Trustors (humans and other objects) live in a partially unpredictable world because of their limited ability to know trustees (surroundings). A trustor is neither in total ignorance nor fully informed concerning a trustee. Under complete ignorance of information, decisions to trust are risky; thus, a transaction of trust should be avoided. Trust becomes meaningless if the trustor has complete information (full rationality) because one can rationally predict before acting. However, in practice, total ignorance and rationality are unrealistic. Humans interact and collaborate in a bounded world where risks are neither fully predictable nor ignored. This situation of partial predictability exposes human beings to a risky world, thus creating a need for trust.

2.3.2. Behavioural Control in Interactive Systems

In recent years, humans, physical robots, bots, and organisations have started to coevolve and interact. These convergent interactions are managed under the security, institutional, and social control approaches. In subsequent paragraphs, attention is drawn to the fact that the term "agent" refers to people, physical robots, bots, and organisations.

a. **Security Control.** Security is a binary control mechanism that attempts to distinguish between agents' contextual behaviour. The security principles offer a sphere of compliant

agents while creating a wall that prevents non-compliant ones. Norms restricting interactions under security control are pre-defined as rules and regulations before being reinforced. Once defined, those rules and regulations are reinforced by scrutinising each agent to verify if it complies. Therefore, security control deals with a binary choice between yes and no, legitimate or illegitimate, acceptance or sanctions [24], authentic or unauthentic, and approval or disapproval. One disadvantage of security controls, particularly in computer systems, is that once an attacker is accepted as genuine and authorised, there is no cap on the privileges it can exercise, rendering it free to carry out (malicious) actions without any extra constraints.

b. **Institutional Control**. Institutional control entails a central authority to monitor, regulate, or enforce the acts taken by agents, and punish those agents who engage in undesired behaviours [25]. For example, police, judicial systems, regulatory bodies, and companies use institutional control to influence the behaviour of individuals and organisations [26] designing for trust in mediated interactions has become a key concern for researchers in human computer interaction (HCI). This form of formal control goes through articulated procedures specifying rewards and punishment. For instance, communication regulatory bodies in various nations and areas monitor online transactions and can testify in court and to the police about cybercrime charges reported.

c. **Social Control**. By enforcing social norms, social control regulates agent interaction in systems. A social norm sanction refers to societal approval or disapproval, which is difficult to determine in advance [27]. Social norms are enforced through social sanctions, which create a range of unpleasant emotional states in those who have violated them [28]. Social control mechanisms don't deny the existence of malicious entities but attempt to avoid interaction with them [29]. In this approach, agents can punish non-desirable behaviours, for instance, by not selecting certain partners [25].

2.3.3. Bases of Trust in Inter-Personal and Business Relationships

Trust is derived from various sources or bases in both personal and business relationships. According to [30], trust can

be based on mechanisms of deterrence, cognition, affection, and calculus, as well as formal and informal institutions. Subsequent paragraphs discuss such bases of trust and how they can apply specifically in cybersecurity incidents (Tab. 2). These bases are adapted from [31].

a. **Calculus-based trust.** Calculus-based trust plays a major role, especially at the beginning of a relationship. As a form of trust-building process, calculus-based trust is founded on: calculating the rewards and costs of committing a transaction, thereby developing confidence that the trustee's behaviour can be predicted, and assessing the trustee's ability to fulfil its promises [32]. Calculus-based trust may be assessed rationally based on credible information sources (reputation, certification) about the trustee. It depends on a rational choice that involves characteristics of interactions founded on economic exchange [33], and deals with factors such as relationship economics and the dynamic capabilities of partners [34]. As the weakest link, humans have to calculate the cost and reward of cooperating based on the level of trust they place in the cybercriminal. Under calculus-based trust, some cybercriminals opt to offer falsified economic benefits, which later turn out to be deception of a victim (computer user).

b. **Deterrence-based trust.** Deterrence occurs when the potential costs of breaking up a relationship outweigh the immediate advantage of acting distrustfully [35]. Deterrence-based trust mechanisms consist of evaluating the advantages and costs of continuing in the relationship, the rewards and costs of cheating on the relationship, and the benefits and costs of quitting the relationship [36].

Table 2. Bases of trust in the human layer of cybersecurity (adapted from [31]).

Basis of trust	Foundation	Description
Process-based trust	Tied to past or expected exchange	Developed based on past or repeated exchanges between cybercriminals and target.
Institution-based trust	Tied to formal social structure, broader societal institutions	Attributes of a person or firm, or an intermediary mechanism shape the possibility for trust to arise.

Basis of trust	Foundation	Description
Deterrence-based trust	Fear of consequences	Behavioural consistency is constrained by the potential costs of discontinuing the relationship.
Competence trust	Based on the partner's competency	An actor predicts others' abilities and expectations of whether they will perform roles competently.
Calculus-based trust	Based on rational choice	Related to the perception of benefit from the relationship.
Relational trust	Tied to repeated interaction	From repeated interaction, the parties obtain information and experience that engenders trust.
Knowledge--based trust	Based on a sufficient understanding of the other party	Prediction of the other party's behaviour based on the history of the relationship.
Identification--based trust	One party has fully internalised the other's preferences	Understanding others' wants. This is the highest level of trust.

Cybercriminals sow fear by threatening users of computer systems to meet falsified demands, which appear to be genuine.

c. **Institutional-based trust.** Institutional trust is tied to formal social structure and broader social institutions. According to [31], the conditions for institutional-based trust are shaped by personal or firm-specific attributes or intermediary mechanisms. Taking advantage of institutional-based trust, cybercriminals impersonate the employees of a particular company, earning the trust of a computer user, who can then be exploited.

d. **Relational trust.** Relational trust refers to the extent to which one feels a personal attachment to the other party and wants to do good for the other party, regardless of egocentric profit motives [37]. The key to relational trust is that one party empathises with the other party and wants to help them for altruistic reasons [37]. Variations of relational trust include normative trust, good will trust, affect-based trust, companion trust, fairness trust, and identification trust [37]. The human psychology of trust is exploited by cyber attackers who understand human perceptions of kindness

and unselfishness well, which exposes cybercrime victims to subsequent consequences.

e. **Identification-based trust.** Identification-based trust involves identification with the other's desires and intentions, i.e. trust exists because one party effectively understands and appreciates the other's wants [38]. This mutual understanding is developed so that each party can effectively act for the other. Identification-based trust is grounded in deep knowledge of the partner's desires and intentions [39]. Identification-based trust can be used to exploit human trust when the trustor and trustee understand each other, as well as when the trustor and trustee have common intentions and desires, e.g. trusting someone to use your electronic gadget. It can also include allowing someone to use your account to access electronic systems, as well as intentionally sharing your credentials with a third party.

f. Knowledge-based trust. Knowledge-based trust is grounded in the other's predictability, or sufficient knowledge that allows the other's behaviour to be anticipated, and relies on information rather than deterrence [38]. Knowledge-based trust develops over time through a track record of interactions that enable both parties to build generalised expectations about each other's behaviour [39]. By being predictable, cybercrime victims are exposed to the actions of the cyber attacker because the cyber attacker knows all the possible means to deceive and manipulate the target, as well as how the target usually responds. Generally, if a person is rationally predictable, that person can be taken advantage of.

2.4. Trust Development Processes

Trust develops from relationship history and subsequent expectations processes. Trust developed from relationship history usually results from the past or previous relationships with people, or other entities or objects [31]. Through relationship history, trust develops based on how parties have previously interacted and the experiences they have gained from one another. When parties have had no previous direct interactions, reference from a third party is usually used to infer the development of trust. Inference is used because, under relationship history, trust develops through interactions with partners that we meet directly or indirectly. Examples of bases of trust that develop from relationship history include

knowledge-based, relational, and process-based. Process-based trust production emphasises that past exchanges, whether through reputation or direct experience, lead to a perception of trust in the counterpart [40].

The second process of trust development involves future expectations. Humans may trust the other party by relying on what they expect to gain after a trust transaction has been performed. Thus, trust formed in this way is usually based on a consideration of the benefits and costs related to a particular relationship [31]. Deterrence-based and calculus-based trust, for example, both rely on future expectations. One party may choose to trust another party after calculating the cost and benefits of an existing relationship. It may also opt to trust because of fear generated by another party.

3. Methodology

The present paper adopts the methodology in [41], developing a theoretical framework that predicts correlations between trust and human behaviours in the cybersecurity layer. In accordance with this methodology, the scope of this paper comprises an analysis of common cyberattacks encountered by users of computer systems, as well as theoretical foundations in trust and cybersecurity. In the former, cybercrime cases reported to police in Tanzania are analysed to indicate how widespread the problem is. Next, a discussion on systems used to control behaviours in interactive systems that fall under face-to-face and computer-mediated environments is presented. Bases of trust that can be used to manipulate human trust are also analysed in detail, and the ways in which trust is employed by cybercriminals to exploit computer users are presented. Generally, most of the discussion is centred on humans as the weakest link in the cybersecurity layer, where human trust is primarily exploited.

Subsequently, the study develops a trust framework to describe how easily human trust can be exploited compared to technical-based controls. This development reveals how deceptive and manipulative attacks on the human psychology of trust go undetected by considering technical and non-technical controls. The study uses data from Tanzania to validate and show the practicality of the framework, which comprise real cases of cybersecurity incidents that were directly observed by the researcher. Secondary data, or cybersecurity incidents reported in the literature originating in Tanzania, are also used.

Additionally, the following are considered during validation and demonstrative application of the trust framework. First, each reported cybercrime incident is explicitly linked to a specific trust formation process. Second, such cybercrime incidents are further linked to bases of trust. This linkage serves to demonstrate how human trust is exploited differently in various circumstances.

Moreover, two issues are taken into consideration throughout the validation and demonstration of application of the trust framework. First, a specific trust formation process is explicitly linked to every cybercrime incident that has been reported. Second, a specific basis of trust is further connected to each cybercrime incident. This connection helps to show how different circumstances lead to different forms of exploitation of human trust.

4. Cybersecurity Trust Framework in the Human Layer

This section details the fundamental structure of trust in the human layer of cybersecurity, which comprises a trustee (cyber attacker) who is regarded as a cybercriminal, and a trustor, usually the end user who is commonly referred to as a cybercrime victim. The cybercriminal and cybercrime victim are the main actors who usually engage in communication.

For a cybercrime incident to occur there must be virtual and occasionally physical interactions between the cybercriminal and victim. Presumably, the victim is protected by technical-based controls, but also policy-based controls, which the victim has to exercise. With those two defences in place, cybercriminals may choose to attack technical protective mechanisms or the human psychology of trust. The latter is the weakest link in the cybersecurity layer. Attacking human psychology of trust is easy to achieve and requires less effort. Most cyberattacks on technological controls can be mitigated by technical countermeasures such as solutions based on encryption, firewall, antivirus, and access control techniques. Insofar as the human layer is excessively exploited, the current framework focuses on exploitation of human trust.

The exchange of exploitative cues between trustor and trustee are hard to detect and prevent by using algorithms inherent in technical-based control systems that protect computer resources (Fig. 2). This is because a criminal communicates as if s(he) is a legitimate user. This may involve forged identities (such as gadgets, or authorised or

unauthorised communication channels) so effectively that computer equipment scrutinising the signals travelling through it uncovers no evidence of susceptibility. These devices are rendered incapable of detecting vulnerability even though they typically perform their protective tasks well based on the functions for which they were built and developed. For example, if a cybercriminal communicates lies via voice or text, the algorithms in those computer devices are unlikely to detect it. Computers rarely detect malevolent intent when a user obtains authorisation and is provided access to systems.

Assume that the cybercriminal wishes to exploit the victim's human psychology of trust (Fig. 2). The cybercriminal must choose the basis of trust to use, depending on whether the cybercriminal and victim have previously interacted. If there has been prior interaction, the cybercriminal will employ bases of trust whose development relies on a relationship history. Otherwise, they will resort to bases of trust that can be developed in anticipation of future expectations.

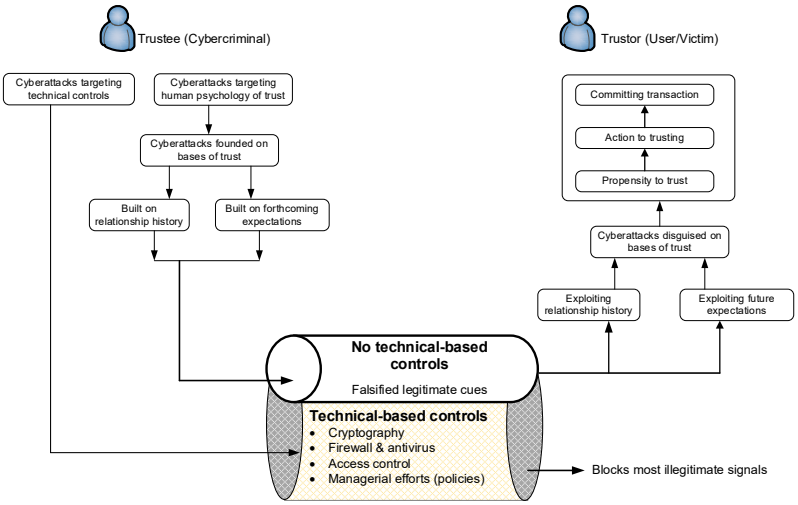

Figure 2. Cybersecurity trust framework in the context of the human layer.

Depending on prevailing circumstances and choice of method or technique, the cybercriminal can use one or multiple bases of trust to deceive and manipulate the victim. Those bases of trust are described fully in subsection 2.3.3. For example, the cyber attacker may use deterrence-based trust to threaten the victim, offering a reward if its demands are met, and punishment if they are not (Table 3). The choice of basis of trust is associated with the trust development process with some using both trust development processes, such

as characteristic-based trust and competence-based trust. Other bases of trust employ one trust development process, for example, calculus-based trust or knowledge-based trust.

Table 3. Bases of trust and their development processes.

Bases of Trust	Trust Development Process	Brief Descriptions
Characteristic-based trust	• Relationship history • Future expectations	Provides background to develop a mutual understanding [31]
Institution-based trust	Relationship history	Attributes of a person or firm, or an intermediary mechanism shape the possibility for trust to arise [31].
Deterrence-based trust	Future expectations	Out of fear, the cybercrime victim meets the cyber attacker's demands, creating future expectations of avoiding harm
Competence trust	• Relationship history • Future expectations	A cybercrime victim develops expectations based on the capability (competence) of the cyber attacker
Calculus-based trust	Future expectations	Weighing the benefit and losses of a relationship
Knowledge-based trust	Relationship history	Behavioural predictability based on available information
Identification-based trust/ Relational trust	Relationship history	A mutual understanding of desires and intentions, as well as the need to feel unselfish and desire to help

The signal/cues communicated by the cybercriminal pass via technical-based solutions undetected. They remain undiscovered because they are (falsified) valid cues in disguise. As a result, the cybercrime victim responds to the cybercriminal, assuming that the signals sent are legitimate.

After the cybercrime attacker gains the victim's trust, it engages the victim in acts that appear innocuous at first. At that stage, the trustor (victim) develops expectations in the fictitious transaction in the hope of obtaining prospects of the assumed agreement.

That stage is known as action to trusting. With that expectation in mind, the victim enters into a transaction after trusting the cybercriminal, performing action(s) that fulfil a promise made between the cybercriminal and victim. Following that transaction, the victim, in accordance with trust standards, compares expectations to the resulting outcome, which may become apparent immediately or later. Furthermore, victims may not realise they have been attacked, depending on the severity of the cyberattack. Some examples include the theft of email passwords where the attacker has no intention of blocking the user's account. In general, the outcome will differ from what the user expected.

5. Case Illustration

The present section provides an illustrative application of the trust framework in the context of the weakest link in cybersecurity. Cybersecurity incidents extracted from real scenarios are presented in Table 4. Some were collected from literature, while others were encountered by the author on various occasions.

In case 1, the cybercriminal uses a mobile phone to deceive users by pretending to be a landlord. The cybercriminal broadcasts information via a Short-Message-Service (SMS) to multiple mobile phone users simultaneously. In that attempt, at least one user may have rented a house from another real landlord, and receives a message concerning rent to be paid. When the SMS is received, the tenant may get confused about whether the sender is the actual landlord or not. The tenant is manipulated further by being directed to pay the rent to a mobile number provided in the SMS.[2] In this situation, the requested amount is expected to be transferred using a mobile money service.

Essentially, the cybercriminal acts as if there is an existing relationship with a victim (landlord-tenant relationship), thereby building trust through a falsified relationship history as a trust development process. The cybercriminal exploits the tenant's trust by employing identification-based trust, the highest psychological tool, to manipulate the tenant into understanding the other side's desire. The cybercriminal has also weaponised relational trust, in which kindness and unselfishness are core components. Overall, the cybercriminal operates on the assumption that there is an agreement on rent payment, taking advantage of the landlord's desire to obtain and the tenant's intention to pay rent.

2 — It should be noted that in some African countries, including Tanzania, mobile phones are used to send and receive money in addition to paying various bills via a service known as mobile money. Mobile phones are used to carry out financial transactions at a country and even regional level.

Table 4. Cybersecurity incidents committed through the exploitation of human trust.

Case No	Incident	Trust Development Process	Bases of Trust Used
1	I'm your landlord. My current number is unreachable. Send the rent through this number +255 (*number withheld*).	(Falsified) relationship history	• Identification-based trust • Relational trust
2	Please get in touch with us as soon as you can; your child is extremely ill. Teacher.	• Future expectations • (Falsified) Relationship history	Deterrence-based trust
3	After unexpectedly collapsing at school, your son was brought to the hospital. Send money right away for medical care.	• Future expectation • (Falsified) Relationship history	Deterrence-based trust
4	Don't call; the phone's speaker is broken; instead, send the money to this number +255 (*number withheld*).	(Falsified) Relationship history	• Identification-based trust • Relational trust
5	This is the Revenue Authority office. Why don't you use an electronic fiscal device (EFD) when conducting business? A Tsh 3 million fine is being sent to you immediately.	Future expectations	• Deterrence-based trust • Calculus-based trust
6	You are speaking with someone from the telecom company (*name withheld*); your monthly bonus is Tsh 400,000 now. Use a different mobile phone so that we can help you obtain the money.	• Future expectations • Relationship history	• Calculus-based trust • Institutional-based trust
7	This is agent (*name withheld*) from telecom company (*name withheld*). Your mobile money account has insufficient funds. Deposit Tsh 500,000 today, then call us back. Otherwise, we are going to close your account.	• Future expectations • (Falsified) Relationship history	• Deterrence-based trust • Calculus-based trust • Institutional-based trust

Case No	Incident	Trust Development Process	Bases of Trust Used
8	I received notification that I had won a customer drawing and was asked to contact a number to learn how to collect my prize. When I called the number, the man instructed me to use 46 as the identification number for prize collection. Then he wanted me to send TSh 60,000 to activate the prize. I sent the money, but when I called the number another time, it was out of service [42].	• Forthcoming expectation • (Falsified) Relationship history	• Institutional-based trust • Calculus-based trust • Characteristics-based trust
9	I received an SMS that appeared to be from M-PESA. The SMS said that I had received TSh 40,000 from a number registered to (*name withheld*). A few minutes after reading the message, someone called and told me he was from Vodacom customer care service. He asked if I had received an SMS that increased my account balance by TSh 40,000. I said I had. Then he asked me to resend the money because it was sent to the wrong account. He told me to send TSh 39,000 to avoid a service charge. When M-PESA replied that the transaction had been successfully completed, I realised my balance had decreased. At that point, I discovered that I had been deceived [42].	Relationship history	• Institutional-based trust • Relational trust

Cybercriminals exploited the tenant's trust psychology because they understand the human perception of kindness and unselfishness. If the tenant cannot sense the deception and use other means to validate whether the received SMS is legitimate, that tenant may end up sending money to a person who is not a real landlord. Such communications pass through digital channels as legitimate cues and are mostly impossible to recognise and filter. Similarly, in case 4, the cybercriminal employs a similar technique to exploit mobile money users.

Cybercriminals also use deterrence-based trust to exploit mobile phone users. In cases 2 and 3, the cybercriminal sends an SMS to targeted parents, informing them that their children are sick. To understand these cases, it should be assumed that some parents send their children to boarding schools. Furthermore, parents are known for their affection and concern for their children; learning that their children are ill can be upsetting and confusing. The cybercriminal (a falsified school teacher or medical doctor) uses deterrence-based trust to introduce a fear that if money is not sent, a child may die from lack of health care. Using deterrence-based trust, the cybercriminal exploits human trust developed through the parent's future expectations of the child's recovery. In addition, the cybercriminal uses institutional-based trust by pretending to be a school teacher, exploiting the parent further to gain trust. This kind of cybercrime employs a common situation in which legitimate teachers and some medical doctors may call parents to obtain additional funds to save a dangerously ill child.

Cases 5, 6, and 7 involve trust building mainly through future expectations and partly through relationship history. In case 5, the cybercriminal communicates via SMS, impersonating an officer of a revenue authority. The falsified officer chooses to create trust with the business owner by setting clear expectations, allowing the owner to believe it is the sole alternative to avoid closure of the business. The business owner is manipulated into believing that if a certain amount is not paid, the revenue authority will close the business. The cybercriminal builds trust through fear (deterrence-based trust) and comparison of the cost and benefit of paying or not paying the falsified fine (calculus-based trust). For cases 6 and 7, the cybercriminal uses mostly future expectations and a (falsified) relationship history to build trust in a mobile money user, relying on calculus-based and institutional-based trust. Through calculus-based trust, a mobile money user compares receiving or losing a bonus (case 6), and making or refusing to make a deposit, and account closure (case 7). Through institutional-based trust, the cybercriminal impersonates an employee of a particular telecom company, gaining more trust from a mobile money user. In case 7, the cybercriminal uses deterrence-based trust to create fear in the mobile money account owner, an agent whose role involves receiving and sending money to mobile money users. The fear is based on the fabricated fact that the account will be cancelled if the owner does not deposit the money. Both cases use relationship history as an additional trust development process. To take advantage of relationship history, cybercriminals impersonate employees of legitimate entities, assuming a legitimate long-term relationship between a mobile money user and a telecom

company. Leveraging that relationship, the mobile money user is further deceived into trusting the cybercriminal.

The last illustration concerns cases 8 and 9, in which both trust development processes are involved. In scenario 8, by setting up future expectations, the cybercriminal communicates that a mobile money user has won a drawing in an attempt to win trust. Falsified winning of the drawing exploits the user's trust as follows: the user compares the benefit and cost (calculus-based trust) of engaging with cybercrime and finally opts to trust because of expectation of winning. The act of trust and committing to sending money is founded on institutional-based trust because the cybercriminal is impersonating an employee of a gambling company. To incorporate relationship history into the trust-building process, the cybercriminal assumes a legitimate long-term relationship that exists between gambling companies and winners. In case 9, the cybercriminal uses relationship history to build trust with a mobile money user. That relationship history is grounded in institutional-based trust because the cybercriminal is impersonating a telecom company employee, exploiting the trust of mobile money users in the company and its employees. To further deepen the trust, the cybercriminal employs relational trust through altruism by asking a mobile money user to return money that was supposedly transferred in error. The mobile money user is exploited by following the cybercriminal's instructions only to find out that their account balance has decreased.

In summary, technical-based controls employed by individuals and organisations rarely detect the above-mentioned techniques of deception and manipulation of human trust since the cues sent to users pass unfiltered via computer-network infrastructure because they are deemed legitimate. Given these limitations, the developed trust framework plays a role in safeguarding users of computer systems.

With respect to the first research question, the present paper argues that cybercriminals exploit human trust based on trust development processes and bases of trust, either creating (falsified) expectations or a relationship history to lure the victim in. Moreover, cybercriminals take advantage of user ignorance of the limitations of technical-based controls. In line with the second research question, the trust framework on exploitation of humans as the weakest link in the cybersecurity layer has many potential benefits and applications. First, the trust framework informs users of computer systems that lies, deception, and manipulation built on human trust can rarely be detected and prevented using technical-based controls. Second, computer system users can identify and stop cybercrime assaults directed at them

by using the trust framework as a guide. Third, people will be less susceptible to cyberattacks if they are aware of the bases of trust that cybercriminals frequently exploit. Finally, computer users will learn and become aware of the way cybercriminals utilise past relationships and future expectations to deceive and carry out cyberattacks.

6. Conclusion

Cybersecurity has become a crucial challenge in this world of digital connectivity because information processing and transfer occurs in cyberspace, which is vulnerable to attacks by many intruders. Recent evidence shows that cyberattacks are increasingly shifting away from technical-based controls to target the human psychology of trust, the weakest link in the cybersecurity layer. Such attacks are linked to how human beings, particularly end users, come to trust cybercriminals. From that viewpoint, the present paper has explored how cybercriminals exploit human trust. In addressing this problem, the paper has established a trust framework to ensure better understanding of how security-based interactions between cybercriminals and victims occur. The framework reveals that trust is a core ingredient in the human – or most vulnerable – layer of cybersecurity. Furthermore, the trust framework indicates that technical-based controls cannot provide effective safeguards to prevent manipulation of the human psychology of trust. Instead, people must protect themselves through greater awareness of cybercrime incidents that are linked to trust. The paper uses real cases to demonstrate the applicability of the trust framework. These scenarios were thoroughly examined, linking them to trust bases and trust development processes. Generally, the sample cases discussed reveal inherent flaws in human trust, which hackers weaponise to deceive computer users.

Despite the extensive discussions presented, this paper has certain limitations and further research may be required to address them. First, this study recognised relationship history and future expectations as trust development processes. In terms of cybersecurity, it is currently unknown which trust development process is more commonly utilised by cybercriminals to exploit human trust. Therefore, future research could investigate common trust development methods employed by cybercriminals. Second, cybercriminals can utilise various bases of trust to deceive and influence computer users. Similarly, greater awareness of which bases of trust cybercriminals employ more often may help organisations and individuals to protect themselves.

References

[1] B. Schneier. (Oct. 15, 2000). "Semantic Attacks: The Third Wave of Network Attacks," *Schneier on Security*. [Online]. Available: https://www.schneier.com/crypto-gram/archives/2000/1015.html#1. [Accessed: Feb. 04, 2023].

[2] URT. (2018). "Crime and traffic incidents:report January-December 2017," *Dar es Salaam*. [Online]. Available: https://www.nbs.go.tz/nbs/takwimu/Crime/Crime_Report_January_to_. December_2017.pdf. [Accessed: Feb. 04, 2023].

[3] Inspector General of Police. (2019). "Takwimu za Hali Ya Uhalifu na Matukio ya Usalama Barabarani Januari –– Desemba 2018," *Dodoma*. [Online]. Available: https://www.nbs.go.tz/index.php/en/census-surveys/crime-statistics. [Accessed: Feb. 04, 2023].

[4] Inspector General of Police. (2020). "Takwimu za Hali Ya Uhalifu na Matukio ya Usalama Barabarani Januari –– Desemba 2019," *Dodoma*. [Online]. Available: https://www.nbs.go.tz/index.php/en/census-surveys/crime-statistics. [Accessed: Feb. 04, 2023].

[5] Inspector General of Police. (2021). "Takwimu za Hali Ya Uhalifu na Matukio ya Usalama Barabarani Januari – Desemba 2020," *Dodoma*. [Online]. Available: https://www.nbs.go.tz/index.php/en/census-surveys/crime-statistics. [Accessed: Feb. 04, 2023].

[6] W. D. Kearney, H. A. Kruger, "Considering the influence of human trust in practical social engineering exercises," in *Proceedings of the ISSA 2014 Conference*, 2014, pp. 1–6, doi: 10.1109/ISSA.2014.6950509.

[7] G. Tejay, G. Klein, "Organizational Cybersecurity Journal editorial introduction," *Organizational Cybersecurity Journal: Practice Process and People*, vol. 1, no. 1, pp. 1–4, 2021, doi: 10.1108/ocj-09-2021-017.

[8] A. Jain, H. Tailang, H. Goswami, S. Dutta, M. S. Sankhla, R. Kumar, "Social Engineering: Hacking a Human Being through Technology," *IOSR Journal of Computing Engineering*, vol. 18, no. 5, pp. 94–100, 2016, doi: 10.9790/0661-18050594100.

[9] DiamondIT. (2022). *The 7 Layers of Cybersecurity*. [Online]. Available: https://www.diamondit.pro/7-layers-of-cybersecurity/. [Accessed: Dec. 03, 2022].

[10] Manhattan Tech Support. (2022). *The seven layers of IT security*. [Online]. Available: https://www.manhattantechsupport.com/. [Accessed Dec. 02, 2022].

[11] D. Henshel, M. G. Cains, B. Hoffman, T. Kelley, "Trust as a Human Factor in Holistic Cyber Security Risk Assessment," *Procedia Manufacturing*, vol. 3, pp. 1117–1124, 2015, doi: 10.1016/j.promfg.2015.07.186.

[12] A. M. Shabut, K. T. Lwin, M. A. Hossain, "Cyber attacks, countermeasures, and protection schemes - A state of the art survey," in *SKIMA 2016–2016 10th International Conference on Software, Knowledge, Information Management and Applications*, 2017, pp. 37–44, doi: 10.1109/SKIMA.2016.7916194.

[13] R. Hanzu-Pazara, G. Raicu, R. Zagan, "The Impact of Human Behaviour on Cyber Security of the Maritime Systems," *Advanced Engineering Forum*, vol. 34, pp. 267–274, 2019, doi: 10.4028/www.scientific.net/aef.34.267.

[14] R. Ottis, P. Lorents, "Cyberspace: Definition and Implications," in *Proceedings of the 5th International Conference on Information Warfare and Security*, 2010.

[15] J. R. C. Nurse, "Cybercrime and You: How Criminals Attack and the Human Factors That They Seek to Exploit," in *The Oxford Handbook of Cyberpsychology*, A. Attrill-Smith, C. Fullwood, M. Keep, D. J. Kuss, Eds., Oxford: Oxford Library of Psychology, Oxford Academic, 2019, pp. 662–690, doi: 10.1093/oxfordhb/9780198812746.013.35.

[16] D. Craigen, N. Diakun-Thibault, R. Purse, "Defining Cybersecurity," *Technology Innovation Management Review*, vol. 4, no. 10, pp. 13–21, 2014, doi: 10.22215/timreview835.

[17] J. R. C. Nurse, S. Creese, M. Goldsmith, K. Lamberts, "Trustworthy and effective communication of cybersecurity risks: A review," in *2011 1st Workshop on Socio-Technical Aspects in Security and Trust (STAST)*, IEEE, Sep. 2011, pp. 60–68, doi: 10.1109/STAST.2011.6059257.

[18] The Citizen. (Dec. 4, 2017). *Cybercrime cases hit 82pc*. [Online]. Available: https://www.thecitizen.co.tz/tanzania/news/business/cybercrime-cases-hit-82pc-2615466. [Accessed: Feb. 04, 2023].

[19] D. Masesa, B. Munyendo, N. Rishad, P. Musuva-Kigen, N. Karumba, *et al.* "Tanzania Cyber Security Report 2016: Achieving Cyber Security Resilience Through Enhancing Visibility and Increasing Awareness," T*anzania Cyber Security Report 2016*, pp. 1–20, 2016. [Online]. Available: http://www.serianu.com/downloads/TanzaniaCyberSecurityReport2016.pdf. [Accessed: Feb. 04, 2023].

[20] AFRIPOL. (2021). African Cyberthreat Assessment *Report: Interpol's Key Insight into Cybercrime in Africa*. [Online]. Available: https://www.interpol.int. [Accessed: Feb. 04, 2023].

[21] TZ-CERT. (2023). "TZ-CERT Honeypots Weekly Report," *Dar es Salaam.* [Online]. Available: https://www.tzcert.go.tz/resources-2/reports/. [Accessed: Feb. 04, 2023].

[22] M. Daudi, *Trust in Sharing Resources in Logistics Collaboration*. Düren: Shaker Verlag GmbH, 2019.

[23] M. Laeequddin, B. S. Sahay, V. Sahay, K. A. Waheed, "Trust building in supply chain partners relationship : an integrated conceptual model," *Journal of Management Development*, vol. 31, no. 6, pp. 550–564, 2012, doi: 10.1108/02621711211230858.

[24] M. Lianos, "Social control after Foucault," *Surveillance & Society*, vol. 1, no. 3, pp. 412–430, 2003.

[25] I. Pinyol, J. Sabater-Mir, "Computational trust and reputation models for open multi-agent systems: A review," *Artifical Intelligence Review*, vol. 40, no. 1, pp. 1–25, 2013, doi: 10.1007/s10462-011-9277-z.

[26] J. Riegelsberger, M. A. Sasse, J. D. McCarthy, "The mechanics of trust: A framework for research and design," *International Journal of Human - Computer Studies*, vol. 62, no. 3, pp. 381–422, 2005, doi: 10.1016/j.ijhcs.2005.01.001.

[27] A. Grizard, L. Vercouter, T. Stratulat, G. Muller, "A peer-to-peer normative system to achieve social order," in: *Coordination, Organizations, Institutions, and Norms in Agent Systems II. COIN 2006. Lecture Notes in Computer Science(),* vol. 4386*, R. Noriega et al. Eds. Berlin, Heidelberg: Springer*, 2007, doi: 10.1007/978-3-540-74459-7_18.

[28] C. R. Sunstein, "Social Norms and Social Rules," *Coarse-Sandor Institute for Law & Economics Working Papers*, vol. 36, 1996.

[29] L. Rasmussen, S. Jansson, "Simulated Social control for Secure Internet Commerce," in *New Security Paradigms Workshop*, C. Meadows, Ed., ACM, 1996. [Online]. Available at: https://www.nspw.org/papers/1996/nspw1996-rasmusson.pdf. [Accessed: Feb. 04, 2023].

[30] A. Capaldo, I. Giannoccaro, "How does trust affect performance in the supply chain? The moderating role of interdependence," *International Journal of Production Economics*, vol. 166, pp. 36–49, 2015, doi: 10.1016/j.ijpe.2015.04.008.

[31] N. P. Nguyen, N. T. Liem, "Inter-Firm Trust Production: Theoretical Perspectives," *International Journal of Business and Management*, vol. 8, no. 7, pp. 46–54, doi: 10.5539/ijbm.v8n7p46.

[32] P. M. Doney, J. P. Cannon, "An Examination of the Nature of Trust in Buyer-Seller Relationships," *Journal of Marketing*, vol. 61, no. April, pp. 35–51, 1997, doi: 10.2307/1251829.

[33] D. M. Rousseau, S. B. Sitkin, R. S. Burt, C. Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of Management Review*, vol. 23, no. 3, pp. 393–404, 1998, doi: 10.5465/AMR.1998.926617.

[34] G. Tejpal, R. K. Garg, A. Sachdeva, "Trust among supply chain partners: A review," *Measuring Business Excellence*, vol. 17, no. 1, pp. 51–71, 2013, doi: 10.1108/13683041311311365.

[35] B. H. Sheppard, D. M. Sherman, "The Grammars of Trust: A Model and General Implications," *The Academy of Management Review*, vol. 23, no. 3, pp. 422–437, 2016, doi:10.2307/259287.

[36] A. F. Salam, L. Iyer, P. Palvia, R. Singh, "Trust in e-commerce," *Communications of the ACM*, vol. 48, no. 2, pp. 72–77, 2005, doi: 10.1145/1042091.1042093.

[37] D. L. Paul, R. R. McDaniel, "A Field Study of the Effect of Interpersonal Trust on Virtual Collaborative Relationship Performance," *MIS Quarterly*, vol. 28, no. 2, pp. 183–227, 2004.

[38] R. J. Lewicki, M. A. Stevenson, R. Lewicki, M. A. Stevenson, "Trust Development in Negotiation: Proposed Actions and a Research Agenda," *Business & Professional Ethics Journal*, vol. 16, no. 1, pp. 99–132, 1997.

[39] J. M. da C. Hernandez, C. C. dos Santos, "Development-based Trust: Proposing and Validating a New Trust Measurement Model for Buyer-Seller Relationships," *Brazilian Administration Review*, vol. 7, no. 2, pp. 172–197, 2010, doi: 10.1590/S1807-76922010000200005.

[40] O. Schilke, G. Wiedenfels, M. Brettel, L. G. Zucker, "Interorganizational trust production contingent on product and performance uncertainty," *Socio-Economic Review*, vol. 1, no. 2, pp. 307–330, 2017, doi: 10.1093/ser/mww003.

[41] E. Jaakkola, "Designing conceptual articles: four approaches," *AMS Review*, vol. 10, no. 1–2, pp. 18–26, 2020, doi: 10.1007/s13162-020-00161-0.

[42] I. H. Bakar. (2016). "Social engineering tactics used in mobile money theft in Tanzania," *The University of Dodoma*. [Online]. Available: http://repository.udom.ac.tz/handle/20.500.12661/1168. [Accessed: Apr. 29, 2023].